

1 **CLAIMS:**

2 1. A method of determining language usage probabilities of a natural
3 language based upon a training corpus, the method comprising:

4 examining a training corpus, wherein such corpus includes phrases parsed
5 in accordance with a set of grammar rules;

6 computing probabilities of usage of combinations of linguistic features
7 based upon empirical tracking of appearances of instances of such combinations in
8 phrases within the training corpus.

9
10 2. A method as recited in claim 1, wherein the combinations of
11 linguistic features comprises:

- 12 • (transition, headword, phrase level, syntactic history, segtype);
13 • (headword, phrase level, syntactic history, segtype);
14 • (modifying headword, transition, headword); and
15 • (transition, headword).
16

17 3. A method as recited in claim 1, wherein the combinations of
18 linguistic features consist of:

- 19 • (transition, headword, phrase level, syntactic history, segtype);
20 • (headword, phrase level, syntactic history, segtype);
21 • (modifying headword, transition, headword); or
22 • (transition, headword).
23
24
25

1 4. A method as recited in claim 1, wherein the computing comprises
2 counting appearances of instances of combinations of linguistic features within the
3 training corpus.

4
5 5. A computer-readable storage medium having computer-executable
6 instructions that, when executed by a computer, performs the method as recited in
7 claim 1.

8
9 6. A method for determining a probability at a node in a parse tree, the
10 method comprising:

11 receiving language-usage probabilities based upon appearances of instances
12 of combinations of linguistic features within a training corpus;

13 calculating the probability at the node based upon linguistic features of the
14 node and the language-usage probabilities.

15
16 7. A method as recited in claim 6, wherein the combinations of
17 linguistic features comprises:

- 18 • (transition, headword, phrase level, syntactic history, segtype);
19 • (headword, phrase level, syntactic history, segtype);
20 • (modifying headword, transition, headword); and
21 • (transition, headword).
- 22
23
24
25

1 8. A method as recited in claim 6, wherein the combinations of
2 linguistic features consist of:

- 3 • (transition, headword, phrase level, syntactic history, segtype);
- 4 • (headword, phrase level, syntactic history, segtype);
- 5 • (modifying headword, transition, headword); or
- 6 • (transition, headword).

7
8 9. A method as recited in claim 6, wherein the calculating comprises
9 using PredParamRule Probability formula to calculate the probability at the node.

10
11 10. A method as recited in claim 6, wherein the calculating comprises
12 using both PredParamRule Probability and SynBigram Probability formulas to
13 calculate the probability at the node.

14
15 11. A method for determining a statistical goodness measure (SGM) of
16 a parse tree representing a parse of a phrase, the parse tree comprising one or more
17 nodes, the method comprising calculating a statistical product of probabilities of
18 each node in the parse tree, wherein the probabilities of each node are determined
19 by the method as recited in claim 6.

20
21 12. A computer-readable storage medium having computer-executable
22 instructions that, when executed by a computer, performs the method as recited in
23 claim 6.

1 13. A method for determining a statistical goodness measure (SGM) of
2 a parse tree representing a parse of a phrase, the parse tree comprising one or more
3 nodes, the method comprising:

4 combining probabilities of each node in the parse tree, wherein the
5 probabilities of each node are determined by the steps comprising:

6 receiving language-usage probabilities based upon appearances of
7 instances of combinations of linguistic features within a training corpus;

8 calculating the probabilities of each node based upon linguistic
9 features of each node and the language-usage probabilities.

10
11 14. A method as recited in claim 13, wherein the combinations of
12 linguistic features comprises:

- 13 • (transition, headword, phrase level, syntactic history, segtype);
14 • (headword, phrase level, syntactic history, segtype);
15 • (modifying headword, transition, headword); and
16 • (transition, headword).

17
18 15. A method as recited in claim 13, wherein the calculating comprises
19 using PredParamRule Probability formula to calculate the probability at the node.

20
21 16. A method as recited in claim 13, wherein the calculating comprises
22 using both PredParamRule Probability and SynBigram Probability formulas to
23 calculate the probability at the node.
24
25

1 17. A method as recited in claim 13, wherein during the combining, the
2 probabilities of each node in the parse tree are combined in a top-down, generative
3 approach.
4

5 18. A method for determining statistical goodness measures (SGMs) of
6 multiple parse trees, each tree representing a syntactically valid parse of a phrase,
7 the method comprising determining a SGM of each parse tree by the method as
8 recited in claim 13.
9

10 19. A method for ranking multiple parse trees, each tree representing a
11 syntactically valid parse of a phrase, the method comprising:
12

13 determining statistical goodness measures (SGMs) of each parse tree by the
14 method as recited in claim 13 to get an SGM values associated with each tree;
15

16 organizing the trees in order of each tree's associated SGM value.
17

18 20. A computer-readable storage medium having computer-executable
19 instructions that, when executed by a computer, performs the method as recited in
20 claim 13.
21

22 21. A method of parsing a phrase to facilitate processing of such phrase
23 by a computer, the method comprising:
24

25 generating at least one parse tree representing a syntactically valid parse of
the phrase, wherein the parse tree has hierarchical nodes;

1 dividing each node into one or more hierarchical phrase levels, wherein the
2 phrase levels at a node represent a set of possible transitions from such node that
3 are allowed by a set of grammar rules.
4

5 **22.** A method as recited in claim 21, wherein the set of possible
6 transitions from each node consists of all possible transitions from such node that
7 are allowed by a set of grammar rules.
8

9
10 **23.** A method as recited in claim 21, wherein the set of possible
11 transitions from each node includes a null transition representing an application of
12 none of the grammar rules.
13

14 **24.** A computer-readable storage medium having computer-executable
15 instructions that, when executed by a computer, performs the method as recited in
16 claim 21.
17

18 **25.** A method of parsing a phrase to facilitate processing of such phrase
19 by a computer, the method comprising:

20 generating at least one parse tree representing a syntactically valid parse of
21 the phrase, wherein the parse tree has hierarchical nodes;

22 calculating a syntactic history for each node.
23
24
25

1 26. A method as recited in claim 25 further comprising storing the
2 syntactic history for each node.

3
4 27. A method as recited in claim 25, wherein the syntactic history may
5 indicate one or more of the following syntactic phenomena:

- 6 • passive verb phrase;
7 • negative polarity;
8 • domodal fronting;
9 • comparative;
10 • imperative;
11 • topicalization of verb object.

12
13 28. A computer-readable storage medium having computer-executable
14 instructions that, when executed by a computer, performs the method as recited in
15 claim 25.

16
17 29. A computer-readable storage medium having computer-executable
18 instructions that, when executed by a computer, determine language usage
19 probabilities of a natural language based upon a training corpus, the method
20 comprising:

21 examining a training corpus, wherein such corpus includes phrases parsed
22 in accordance with a set of grammar rules;
23
24
25

1 computing probabilities of usage of combinations of linguistic features
2 based upon empirical tracking of appearances of instances of such combinations in
3 phrases within the training corpus.
4

5 **30.** A computer-readable storage medium having computer-executable
6 instructions that, when executed by a computer, perform a method to determine a
7 probability at a node in a parse tree, the method comprising:

8 receiving language-usage probabilities based upon appearances of instances
9 of combinations of linguistic features within a training corpus;

10 calculating the probability at the node based upon linguistic features of the
11 node and the language-usage probabilities.
12

13 **31.** A computer-readable storage medium having computer-executable
14 instructions that, when executed by a computer, perform a method to determine a
15 statistical goodness measure (SGM) of a parse tree representing a parse of a
16 phrase, the parse tree comprising one or more nodes, the method comprising:

17 combining probabilities of each node in the parse tree, wherein the
18 probabilities of each node are determined by the steps comprising:

19 receiving language-usage probabilities based upon appearances of
20 instances of combinations of linguistic features within a training corpus;

21 calculating the probabilities of each node based upon linguistic
22 features of each node and the language-usage probabilities.
23
24
25

1 **32.** A computer-readable storage medium having computer-executable
2 instructions that, when executed by a computer, perform a method to rank multiple
3 parse trees, each tree representing a syntactically valid parse of a phrase, the
4 method comprising:

5 generating at least one parse tree representing a syntactically valid parse of
6 the phrase, wherein the parse tree has hierarchical nodes;

7 dividing each node into one or more hierarchical phrase levels, wherein the
8 phrase levels at a node represent a set of possible transitions from such node that
9 are allowed by a set of grammar rules.

10
11 **33.** A computer-readable storage medium having computer-executable
12 instructions that, when executed by a computer, perform a method to parse a
13 phrase, the method comprising:

14 generating at least one parse tree representing a syntactically valid parse of
15 the phrase, wherein the parse tree has hierarchical nodes;

16 calculating a syntactic history for each node.

17
18 **34.** An apparatus comprising:

19 a processor;

20 a natural-language-usage probability determiner executable on the
21 processor to:

22 examine a training corpus, wherein such corpus includes phrases
23 parsed in accordance with a set of grammar rules;

1 compute probabilities of usage of combinations of linguistic features
2 based upon empirical tracking of appearances of instances of such
3 combinations in phrases within the training corpus.
4

5 **35.** An apparatus as recited in claim 34, wherein the combinations of
6 linguistic features comprises:

- 7 • (transition, headword, phrase level, syntactic history, segtype);
- 8 • (headword, phrase level, syntactic history, segtype);
- 9 • (modifying headword, transition, headword); and
- 10 • (transition, headword).

11
12 **36.** An apparatus as recited in claim 34, wherein the combinations of
13 linguistic features consist of:

- 14 • (transition, headword, phrase level, syntactic history, segtype);
- 15 • (headword, phrase level, syntactic history, segtype);
- 16 • (modifying headword, transition, headword); or
- 17 • (transition, headword).

1
2 **37.** An apparatus comprising:
3 a processor;
4 a natural-language-usage probability determiner executable on the
5 processor to:

6 receive language-usage probabilities based upon appearances of
7 instances of combinations of linguistic features within a training corpus;
8 calculate a probability at a node in a parse tree based upon linguistic
9 features of the node and the language-usage probabilities.
10

11 **38.** An apparatus as recited in claim 37, wherein the combinations of
12 linguistic features comprises:

- 13 • (transition, headword, phrase level, syntactic history, segtype);
14 • (headword, phrase level, syntactic history, segtype);
15 • (modifying headword, transition, headword); and
16 • (transition, headword).
17

18 **39.** An apparatus as recited in claim 37, wherein the combinations of
19 linguistic features consist of:

- 20 • (transition, headword, phrase level, syntactic history, segtype);
21 • (headword, phrase level, syntactic history, segtype);
22 • (modifying headword, transition, headword); or
23 • (transition, headword).
24
25

1
2 **40.** An apparatus as recited in claim 37, wherein the determiner
3 calculates the probability at the node by using PredParamRule Probability
4 formula.
5

6 **41.** An apparatus as recited in claim 37, wherein the determiner
7 calculates the probability at the node by using both PredParamRule Probability
8 and SynBigram Probability formulas.
9

10 **42.** An apparatus comprising:

11 a processor;

12 a natural-language-usage parser executable on the processor to:

13 generate at least one parse tree representing a syntactically valid
14 parse of the phrase, wherein the parse tree has hierarchical nodes;

15 divide each node into one or more hierarchical phrase levels,
16 wherein the phrase levels at a node represent a set of possible transitions
17 from such node that are allowed by a set of grammar rules.
18

19 **43.** An apparatus as recited in claim 42, wherein the set of possible
20 transitions from each node includes a null transition representing an application of
21 none of the grammar rules.
22
23
24
25

1 **44.** An apparatus comprising:

2 a processor;

3 a natural-language-usage parser executable on the processor to:

4 generating at least one parse tree representing a syntactically valid
5 parse of the phrase, wherein the parse tree has hierarchical nodes;

6 calculating a syntactic history for each node.

7
8 **45.** An apparatus as recited in claim 44, wherein the syntactic history
9 may indicate one or more of the following syntactic phenomena:

- 10 • passive verb phrase;
- 11 • negative polarity;
- 12 • domodal fronting;
- 13 • comparative;
- 14 • imperative;
- 15 • topicalization of verb object.

16
17 **46.** A natural-language-usage probability determiner comprising:

18 data-acquisition device for receiving language-usage probabilities based
19 upon appearances of instances of combinations of linguistic features within a
20 training corpus;

21 probability calculator for calculating a probability at a node of a parse tree
22 based upon linguistic features of the node and the language-usage probabilities.

23
24
25

1 **47.** A data structure for use with a computer having a processor and a
2 memory, said structure comprising:

3 a corpus comprising one or more phrases in a natural language;

4 parse trees having hierarchical nodes, each tree representing at least one
5 syntactically valid parse of each phrase in a subset of the corpus;

6 wherein each node has one or more hierarchical phrase levels, wherein the
7 phrase levels at a node represent a set of possible transitions from such node that
8 are allowed by a set of grammar rules.

9
10 **48.** The structure as recited in claim 47, wherein the subset of the corpus
11 includes all phrases in the corpus.

12
13 **49.** A data structure for use with a computer having a processor and a
14 memory, said structure comprising:

15 a corpus comprising one or more phrases in a natural language;

16 parse trees having hierarchical nodes, each tree representing at least one
17 syntactically valid parse of each phrase in a subset of the corpus;

18 wherein one or more nodes have a syntactic history associated therewith.

19
20 **50.** The structure as recited in claim 49, wherein the subset of the corpus
21 includes all phrases in the corpus.

1 **51.** A data structure for use with a computer having a processor and a
2 memory, said structure comprising:

3 a corpus comprising one or more phrases in a natural language;
4 parse trees having hierarchical nodes, each tree representing at least one
5 syntactically valid parse of each phrase in a subset of the corpus;

6 wherein each node as an associated probability, wherein the associated
7 probability of a node is based upon linguistic features of such node and language-
8 usage probabilities derived from appearances of instances of combinations of
9 linguistic features within a training corpus.

10
11 **52.** A method as recited in claim 51, wherein the combinations of
12 linguistic features comprises:

- 13 • (transition, headword, phrase level, syntactic history, segtype);
14 • (headword, phrase level, syntactic history, segtype);
15 • (modifying headword, transition, headword); and
16 • (transition, headword).

17
18 **53.** A method as recited in claim 51, wherein PredParamRule
19 Probability formula is used to calculate a probability associated with a node.

20
21 **54.** A method as recited in claim 51, wherein both PredParamRule
22 Probability and SynBigram Probability formulas are used to calculate a probability
23 associated with a node.
24
25

1
2 **55.** The structure as recited in claim 51, wherein the subset of the corpus
3 includes all phrases in the corpus.
4

5 **56.** A program module for execution on a computing operating
6 environment having a memory, the module comprising:

7 a natural language phrase parser configured to generate one or more
8 syntactically valid parses for a phrase, each parse may be represented by a parse
9 tree having hierarchical nodes;

10 a parse ranker configured to calculate a SGM for each parse of a phrase and
11 to rank the parses to indicate a most probable parse;

12 wherein the parse ranker comprises:

13 data-acquisition device for receiving language-usage probabilities
14 based upon appearances of instances of combinations of linguistic features
15 within a training corpus;

16 probability calculator for calculating a probability at a node of a
17 parse tree based upon linguistic features of the node and the language-usage
18 probabilities.
19

20 **57.** A natural language processing system comprising a program module
21 as recited in claim 56.
22
23
24
25

1 **58.** A grammar checking system comprising a program module as
2 recited in claim 56.

3
4 **59.** A speech processing system comprising a program module as
5 recited in claim 56.

6
7 **60.** A database query processing system comprising a program module
8 as recited in claim 56.

9
10 **61.** An operating system comprising a program module as recited in
11 claim 56.
12
13
14
15
16
17
18
19
20
21
22
23
24
25